Universidad Autónoma de Madrid

Escuela Politécnica Superior

Doble Grado en Ingeniería Informática y Matemáticas

TRABAJO DE FIN DE GRADO

MODELADO DE DATOS DE TURISMO CON REDES NEURONALES

Autor: Daniel Laorden Martín
Tutor: David Renato Domínguez Carreta

Junio 2018

MODELADO DE DATOS DE TURISMO CON REDES NEURONALES

Autor: Daniel Laorden Martín
Tutor: David Renato Domínguez Carreta

Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid

Junio 2018

Agradecimientos

Un viaje de mil millas comienza con el primer paso.
Lao Tse

Muchas gracias a David por su consejo y por haberse amoldado a mis circunstancias.

Todo

Abstract

Tourism is a mainstay of the Spanish economy, to such an extent that it will represent approximately 15 % of GDP, and generates more than two and a half million jobs. Besides, it is directly connected with different industries such as hospitality, leisure or transport. With these premises, understanding the shape and behavior of the tourism industry is at the forefront of many companies and public organizations in Spain.

Therefore, the goal of this thesis is to be introduced in the study of tourism from the point of view of Automatic learning. A variety of techniques are utilized (regression, clusterization and classification) against data from INE about arrivals and tourist spending, with aims of forecasting demand and categorizing incoming tourists in different segments.

Various classifiers have been used for demand forecasting (multilayer perceptron, support vector machines) depending on the time series, as their cycles and shape vary (between different autonomous communities, for example). The thesis contains forecasting for the next 3 and 5 years by origin country, and up to 3 years for autonomous community.

For segmentation, different clustering algorithms have been used both for arrival data (2013 - 2015) and for spending data (2013), from which we obtain five and nine clusters respectively.

Key words — Automatic learning, Big Data, Classifiers, Forecasting, Neural networks, Segmentation, Society, Time series, Tourism

Resumen

El turismo es un pilar clave de la economía española. Tanto, como que representará en 2018 aproximadamente un 15 % del PIB y genera más de dos millones y medio de puestos de trabajo. Además está directamente relacionado con diferentes industrias, cómo la hostelería, el ocio o el transporte. Con estas premisas, estudiar y entender la forma y el comportamiento del turismo están en el punto de mira de empresas y organizaciones públicas españolas.

Por tanto, el objetivo de este trabajo es introducirse en ese estudio desde el ángulo del aprendizaje automático. Se aplican una serie de técnicas (regresión, clusterización y clasificación) a datos del Instituto Nacional de Estadística sobre llegadas y gasto turístico, con la intención de predecir demanda, y categorizar los turistas entrantes en distintos segmentos.

Se han utilizado distintos clasificadores para la previsión de demanda (perceptrón multicapa, máquinas de vectores de soporte), en función del tipo de serie temporal (ya que se han detectado ciclos y patrones en función de la Comunidad Autónoma, por ejemplo), y el trabajo contiene predicción del numero de turistas a 3 y 5 años en función de su país de origen, y hasta 3 años por Comunidad Autónoma.

Para la segmentación se ha utilizado clustering tanto en la información de entrada de turistas desde 2013 hasta 2015 como en la de gasto del año 2013, de donde se obtienen cinco y nueve clusters respectivamente.

Palabras clave — Aprendizaje automático, Big Data, Clasificadores, Redes neuronales, Segmentación, Series temporales, Sociedad, Turismo, Predicción de tendencias

Índice general

Índice de figuras

1 Introducción

1.1. Motivación

Para algunos países, el turismo es un pilar clave en ámbitos como la economía, la educación o la cultura. Con la globalización, el aumento de poder adquisitivo en el mundo y la bajada de precios en transporte, se ha disparado y se prevé que continúe en esta dirección.

En España, estamos ante nuestra mayor industria, aportando 110.000 millones de euros a la economía, más de un 11 % del PIB en 2015 [1]. Genera miles de puestos de trabajo al año, y fue clave en nuestra recuperación económica [2]. Su importancia hace que tengamos que destinar recursos hacia nuevas vías de mejora, con diferentes fines: desde atraer nuevos turistas, hasta adaptar su estancia en el país para que no haya conflicto con la población local.

Algunos de nuestros esfuerzos son un éxito, hemos superado a Estados Unidos como segundo país en número de visitantes al año, y probablemente obtengamos la primera posición pronto [3]. Sin embargo, distintas regiones de España exponen su rechazo hacia determinados perfiles turísticos, y el desarrollo de nuevas tecnologías adaptadas al turismo (SmartCities, por ejemplo) suponen nuevos retos ante los que hay que continuar evolucionando.

Con estas premisas, y teniendo en cuenta la amplísima cantidad de información disponible sobre las llegadas a cualquier país de Europa, es vital su análisis. A este respecto se han elaborado diferentes métricas que proporcionan información sobre la competitividad del turismo en un determinado lugar. Además, existen correlaciones entre datos del viajero y preferencias de destino: la edad o el país de procedencia son buenos estimadores de viaje.

Por tanto, cualquier empresa o gobierno que se precie estudiarán en detalle esta información para generar tendencias, realizar predicciones y adquirir conocimiento que proporcione ventajas competitivas.

1.2. Definición del proyecto

1.2.1. Objetivos y alcance

Este trabajo tiene dos objetivos principales: el primero, la elaboración de un informe que englobe los puntos más importantes del turismo en España (demografía, edades, destinos más visitados, impacto económico y social) durante los últimos años, incluyendo predicciones para el futuro. El segundo objetivo es la búsqueda y utilización de las técnicas de aprendizaje automático más apropiadas para aplicar a datos en el ámbito turístico.

Para llevar a cabo estos objetivos se deberán resolver las siguientes cuestiones:

- Primero, la obtención de datos. Se requiere toda la información posible, por lo que habrá que buscar en bases de datos especializadas, agencias de turismo a nivel ciudad / país / internacionales, y cualquier otra fuente de información. Con estos datos en bruto probablemente haya que crear una clasificación propia de manera que se mantengan únicamente los indicadores más importantes para el trabajo.
- A continuación entraría la parte más importante del trabajo, el análisis de los datos obtenidos en el paso anterior para obtener conclusiones. Se utilizará la librería Weka, especializada en big data y algoritmos de aprendizaje automático. Los clasificadores serán personalizados para asegurarnos de que se adecúan perfectamente a este proyecto. En este apartado se busca obtener conclusiones puntuales, agrupamiento mediante clustering de países que nos visitan, utilizando diferentes métricas, y predicciones a futuro. Todo este análisis estará validado estadísticamente, teniendo en todo momento una horquilla con el error controlado.

El alcance abarca todas las inferencias que se puedan obtener con los datos libres proporcionados por agencias en los ámbitos mencionados anteriormente. Se ha dejado fuera toda la información obtenible por vía privada: datos transaccionales, datos telefónicos, encuestas privadas por parte de empresas dedicadas al turismo.

1.2.2. Tecnología y herramientas empleadas

- Java: Dado que la API de Weka está en Java, toda la parte del proyecto con los clasificadores está programada en este lenguaje. Se ha utilizado bastante del código ya existente, y se han variado ciertos parámetros para optimizar el resultado final.
- R: Se ha utilizado R para hacer mapas geográficos con las predicciones de demanda.
- SQL: La base de datos está creada en MySQL, se ha utilizado la herramienta oficial (MySQL Workbench) y se han creado diferentes consultas para generar los ficheros ARFF.
- Excel: Los datos que no forman parte de la BD han sido descargados en formato csv, y han sido preprocesados en Excel.
- Python: El código que calcula los Exponentes de Lyapunov está desarrollado en Python.

1.2.3. Metodología

A la hora de desarrollar cada módulo se ha utilizado una metodología incremental. Iterativamente se encontraba la solución a un problema planteado a partir de los datos disponibles, y se pensaba el siguiente problema a resolver. Por tanto, cronológicamente desde el inicio del trabajo:

La primera parte fue la búsqueda de datos en bruto. Se buscó en diferentes portales y repositorios de datos, y se ha optado por utilizar los datos del Instituto Nacional de Estadística (INE), previamente de Turespaña. Estos datos se componen de varias series temporales con información sobre entradas a España y preferencias de gasto de turistas. Se extrajeron de dos formas, una primera aproximación fue extraer especificamente las series "limpias", y estas series se utilizaron para la parte de la demanda. Para clusterización se optó por extraer todo el bloque de "Microdatos", con más nivel de detalle y más variables para elegir.

Estos ficheros de datos en bruto pasaron por programas de procesamiento de texto en Java, y luego entraron en una base de datos a partir de la cual se generan los ARFF, en función de una selección de variables hecha previamente con las consideradas más interesantes (donde interesante ha significado dar más información junto a ser fácilmente procesable por los algoritmos).

Se hicieron pruebas con diferentes algoritmos tanto de regresión cómo de clusterización y se tomaron los más óptimos en tanto a las métricas de error consideradas, y tras ellas se han creado las gráficas de predicción de demanda y la segmentación de turistas. La primera versión fue una serie de gráficas con las series temporales y su predicción para comunidades autónomas, después se añadieron los mapas de comunidades autónomas y en una tercera iteración se añadió la parte que estudia la demanda por país de orígen.

Por último, una vez que todos estos módulos se completaron, se ha añadido toda la información a esta memoria y se ha desarrollado el apartado de conclusiones resumiendo todo lo aprendido en el proceso.

1.3. Estructura del documento

Esta memoria está dividida en cinco secciones:

- La primera sección describe la motivación de este trabajo, e introduce el alcance, las metas, que tecnología se ha utilizado y la metodología.
- La segunda sección contiene el estado del arte. Se describe qué es el turismo, una perspectiva histórica del turismo (especialmente para España) y de distintas métricas y cuantificaciones en él. Por otro lado, se introducen las series temporales y teoría sobre su estudio. Por último se discute la problemática y retos futuros del turismo.
- La tercera sección contiene información sobre cómo se ha desarrollado (diseño e implementación) todo el proceso desde encontrar datos hasta el punto en el que ya se puede empezar con los distintos estudios. Además, se describen los algoritmos y métricas utilizados con mayor detalle.
- En la cuarta sección se expone el análisis realizado. Hay una parte de predicción de tendencias para estimar la demanda turística de los próximos años, y una parte de segmentación de turistas desde dos ángulos diferentes, los datos de demanda y los datos de gasto.
- La quinta y última sección comprende las conclusiones de este trabajo, tanto las extraídas puramente por los datos como las conclusiones sobre lo que ha sido realizar este trabajo y lo aprendido.

2

Estado del Arte

2.1. Introducción. El turismo en la actualidad

La Organización Mundial del Turismo define el turismo cómo "las actividades que realizan las personas durante sus viajes y estancias en lugares distintos a su entorno habitual por un período de tiempo consecutivo inferior a un año, con fines de ocio, negocios u otros". Es una actividad que realizamos desde tiempos inmemoriables, sin embargo, conforme la sociedad ha ido avanzando a lo largo del tiempo, las maneras de hacer turismo han ido cambiando a la par.

En los últimos 40 años se ha vivido una revolución en la dinámica del turismo mundial: la globalización ha multiplicado el turismo internacional (organismos como la Unión Europea fomentan la diversidad cultural), y la mejora de los medios de transporte permite recorrer mayores distancias a menor coste. Específicamente en España, donde hemos multiplicado nuestro número de turistas enormemente, se ha pasado de un modelo turístico uniforme, a lo largo y ancho de la península, a un modelo nuclear en el que unas pocas urbes están masificadas (y en algunos casos, generando aversión en la población local).

Además estos cambios han forzado a los negocios turísticos a modificar sus estrategias. Ahora los turístas son exigentes, buscan lugares únicos, con actividades diferenciadoras, lo que fuerza al sector a ser más competitivo e innovador. En este aspecto España está especialmente bien posicionada, de hecho, el Foro Económico Mundial sitúa España en el puesto número 1 mundial en competitividad turística [4], por primera vez, gracias a sus avances en el desarrollo digital del turismo.

Pero, ¿cuáles son las claves para mantener este éxito de manera sostenida en el futuro? Deloitte, en su último informe sobre la evolución del turismo en España [5], señala tres principales, de las cuales dos tienen una especial relevancia para nosotros: Estrategia digital (con un peso del 40 %) y Tecnologías Exponenciales: uso de Big Data o Inteligencia Artificial (35 %).

2.2. Indicadores principales

El turismo tiene un gran impacto tanto en la cultura como en el entorno físico en el que se desarolla, por tanto, una gran rama de su análisis está orientada a la sostenibilidad. Existen una serie de manuales y políticas establecidas por la UNWTO, orientados al control de áreas cómo:

- Preservación de patrimonio histórico y arquitectura tradicional
- Gestión de resíduos y medio ambiente
- Uso de la agricultura local en el turismo
- Incremento de precios del suelo e inmobiliario
- Población que busca trabajos cualificados
- Impacto en la comunidad local

Para estudiar los indicadores relacionados con estas categorías, se recolectan datos de cuatro maneras principales: información pública por parte de gobiernos ú organismos internacionales, mediante encuestas, procesos interactivos o métodos no intrusivos (p.ej. fotografías aéreas).

Como ejemplo práctico de un informe actual sobre la sostenibilidad en el turismo tenemos el análisis de Travel Foundation and TUI Group y PWC [6] a ocho hoteles en Chipre, utilizando una métrica propia, el TIMM, que agrega una media ponderada de todos los impactos (social, económico, etc...) y devuelve una cifra exacta.

Además de la sostenibilidad, la competitividad y el beneficio económico son las otras dos ramas principales. Para medirlas, se utilizan indicadores cómo el número de estancias (o porcentaje de ocupación) por unidad de tiempo, el gasto medio por persona, la duración media de viaje por persona, el tráfico de viajeros por tierra/mar/aire o el número de visitas a un determinado lugar.

Un ejemplo de informe sobre competitividad es "The Travel & Tourism Competitiveness Report", creado anualmente por el World Economic Forum [7], en el que expertos en viajes y turismo evalúan en detalle los factores y políticas más importantes para obtener el mayor beneficio económico de la forma más eficaz para un país.

2.3. Estudio de series temporales: métodos y técnicas

Una serie temporal es una colección de observaciones o muestras ordenadas cronológicamente en una distribución bidimensional, en la que normalmente el eje horizontal es el tiempo y el vertical la variable que estemos estudiando. Históricamente, se han encontrado series temporales que datan del S. XI, aparentemente sobre los movimientos del sol y varios planetas (Tufte 1983).

El estudio como tal se desarrolló espectacularmente durante el S. XX. Se desarrolló toda la teoría sobre procesos estocásticos que precedió a las primeras aplicaciones de regresión automática sobre series temporales, en las décadas de 1920 y 1930 (G. U Yule, J. Walker). [8][9] En los años 70 salió a la luz el libro "Time Series Analysis"[10], de G. E. P. Box and G. M. Jenkins, que mejoró significativamente las técnicas de predicción y las aproximaciones. Se pasó de utilizar medias móviles a modelos sofisticados, primero ARMA y luego ARCH y derivados.

Además de los métodos tradicionales, desde 1990 en adelante se empiezan a utilizar técnicas de aprendizaje automático [11] para modelar las series y minimizar el error. Principalmente son utilizadas redes neuronales (LSTN, RNN) y actualmente otras técnicas como algoritmos genéticos o evolutivos se utilizan para complementar dichas redes.

2.4. Problemática y retos futuros

Uno de los retos principales es la obtención de datos. Grandes corporaciones recurren a empresas especializadas que realizan encuestas orientadas a resolver un problema en particular, pero entidades con menos recursos (por ejemplo, para la elaboración de este trabajo) no disponen de los mismos medios, lo que les deja en desventaja. Además, la monopolización de ciertos datos es una realidad cada vez más común, dónde compañías telefónicas poseen toda la información de sus clientes, o bancos poseen toda la información transaccional. Ambos son claves para hacer determinados estudios sobre turismo, que no pueden hacerse con los datos tradicionales que proporcione un gobierno o la ONU.

Otro reto muy importante en el sector turismo es la innovación. Prácticamente todos hacen operaciones comunes cómo predicción de demanda ú ocupación, actualmente el valor añadido está en hacer herramientas fuera de lo común, como por ejemplo la aplicación BBVA Tourism que muestra el gasto en España en 2014, a un gran nivel de detalle [12]. Convenientemente, utiliza la base de datos de transacciones de BBVA, por lo que comprobamos que si queremos innovación, hace falta cierta liberalización de los datos.

3 Diseño

3.1. Tratamiento de datos

Para elaborar las base de datos se han utilizado variables de tres tipos:

- Temporal: cada registro tiene un campo con el año y el mes de la muestra
- Geográfico:
  - Comunidad autónoma: este campo marca la comunidad autónoma visitada por la muestra
  - País de residencia: este campo marca la procedencia de los turistas de esta muestra
- Económico:
  - Tipo de viaje: ocio, negocios ú otros motivos
  - Gasto diario por persona: marca la cantidad gastada al día por cada turista
  - Alojamiento: dónde se ha alojado (hotel, camping, alquiler...)
  - Via de entrada: que medio de transporte ha utilizado para entrar
  - Tipo de viajero: distingue entre turista o excursionista, y entre residente o no residente
  - Número de noches por persona: cuántos días ha pernoctado (de media) un turista en su destino
  - Número total de personas que han viajado durante el mes de la muestra

Además para la base de datos de Egatur se utilizan variables ternarias que marcan los gustos de los turistas (1 = realizan actividad gratis, 0 = no la realizan, 6 = la realizan pagando), por ejemplo, asistió a espectáctulos deportivos, alquiló coche, contrató pack de servicios turísticos, visitó restaurantes.

```
@DATA
12,3,9,6,1,1,1,1,1,1,1,6,1,6,6,6,1,1,6,1,2,2250
13,5,6,0,0,0,0,0,0,0,0,6,1,6,1,1,6,6,1,1,1,2000
4,9,6,0,0,0,0,0,0,0,0,6,6,6,1,6,1,6,1,1,2,210
4,9,6,0,0,0,0,0,0,0,0,6,6,6,1,6,1,6,1,1,2,210
15,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1250
```

Figura 3.1: Muestra de la base de datos de Egatur

Para empezar, se han tomado las series de datos del Instituto de Estudios Turísticos (hasta 2015) y INE (hasta actualidad), y dado que habían bastantes cruces de datos que no aparecían en estas series, se han utilizado los microdatos de la base de datos FRONTUR. Con estos microdatos se ha creado una base de datos desde la cual se han generado los ficheros ARFF necesarios para WEKA, y los CSV necesarios para los mapas en R.

Durante esta parte han surgido una serie de problemas que ha habido que ir solucionando:

El primer problema es que cada fichero de microdatos contiene información anual en bruto, para 2013, por ejemplo, había más de 4 millones de filas. Para solucionarlo se ha optado por crear la base de datos y eliminar variables que se han considerado poco importantes para este estudio, cómo la autopista/aeropuerto exactos que utilizó el turista para entrar.

Una vez creada la base de datos, el siguiente problema fue que no coincidía a priori el numero total de turistas con el oficial del INE. Para 2013 el dato oficial son 60 millones de turistas y en la base de datos se obtenía 160 millones. Tras estudiar los datos se vió que había muchos duplicados, tanto por ID como por toda la fila. Se limpió la base de datos de duplicados y el numero de turistas obtenido era de unos 20 millones, menor al oficial. Al final, a base de prueba y error, la solución fue filtrar los datos por el parámetro "Motivos del viaje", y tomar sólo un tipo de turistas.

Para generar los ficheros ARFF se han hecho varias consultas en SQL que exportan a CSV, y luego se ha manipulado los ficheros para convertirlos en ARFF. Las primeras pruebas con WEKA dieron muy malos resultados ya que, tal y como vienen las líneas de datos de FRONTUR, cada línea marca cuántos turístas del total cumplen unas determinadas condiciones (frecuencia). Ningun algoritmo de WEKA estaba condicionado para analizar frecuencias así que había dos opciones, modificar los algoritmos o modificar el fichero de datos. Se optó por la segunda opción y se ha creado un pequeño programa en Java que lee los datos con número de viajeros por fila en CSV y devuelve datos sin este ultimo atributo, añadiendo una fila por cada viajero con sus características.

3.2. Análisis de los algoritmos

Se han utilizado algoritmos de regresión para hacer predicción, directamente sobre las series temporales de demanda. A continuación se utilizan algoritmos de clustering para segmentar los datos en distintos perfiles, y, por último, a partir de los clusters formados, se utilizan algoritmos de clasificación para comprobar si se puede predecir a que segmento pertenece un turista cualquiera a partir de sus datos.

De esta manera englobamos en este trabajo una aplicación de cada grupo de técnicas de minería de datos [13].

3.2.1. Algoritmos de regresión

Estos algoritmos se basan en estudiar la relación entre los resultados de una serie de variables independientes para determinar una variable dependiente [14]. En este caso, aplicados a series temporales, dada una serie de observaciones en el tiempo $S_1, ..., S_N$ el modelo general se describe como [15]:

$$s_t = g(t) + \psi_t$$

donde $g(t)$ es una función determinista dependiente del tiempo, y ψ_t es el ruido de la serie, y suele ser probabilística o está basada en una combinación de los errores de la serie.

Regresión lineal

Es el algoritmo de regresión más básico. Se basa en ajustar una ecuación lineal a partir de los datos que se introduzcan. Para su implementación se ha utilizado WEKA con los parámetros de serie.

Optimización Mínima Secuencial (SMO)

Este algoritmo [16] es un algoritmo de entrenamiento de SVM (Support Vector Machine [17]). Puede utilizarse tanto para regresión como para clasificación, y han sido probados con buen resultado en problemas de reconocimiento de caracteres [18] o detección de caras [19], entre otros usos.

Un algoritmo basado en SVM separa el conjunto de datos entre un grupo negativo y uno positivo (el hiperplano que separa ambas muestras es el Support Vector Machine) y define el margen como la máxima distancia de ambos conjuntos al hiperplano. La particularidad de SMO es que utiliza multiplicadores de Lagrange para optimizar el cálculo de este hiperplano, y por tanto tiene un rendimiento mejor al de métodos de entrenamiento previos.

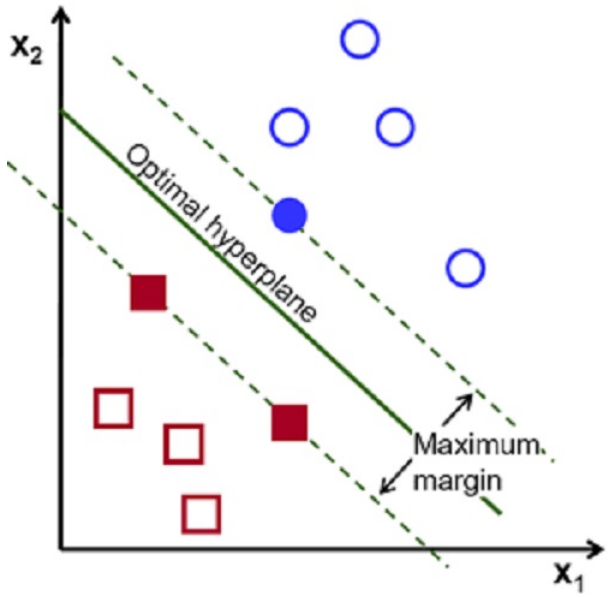


Figura 3.2: Funcionamiento de un SVM

Se ha utilizado la implementación SMOReg de WEKA, que contiene un método de optimización extra para regresión [20].

Preceptrón Multicapa

Este algoritmo está basado en el uso de redes neuronales para la regresión. Genera una red neuronal formada por varias capas. Las redes neuronales tienen los mejores resultados si son bien entrenadas para el problema a resolver, pero si el entrenamiento no es ideal los resultados son imprecisos. Es utilizado, por ejemplo, en meteorología [21].

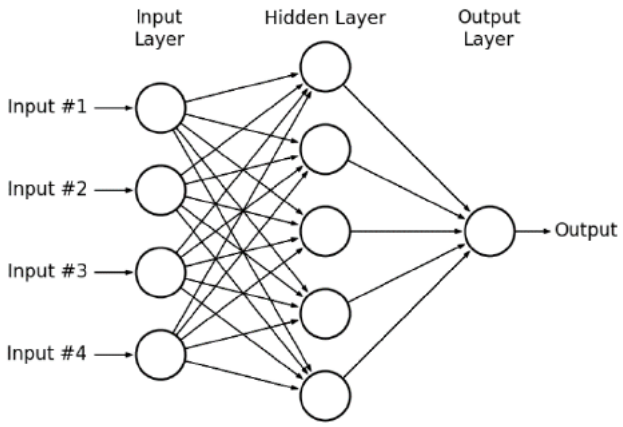


Figura 3.3: Estructura del Preceptrón Multicapa

Se ha implementado utilizando MultilayerPreceptrón en WEKA, que permite modificar el número de neuronas en la capa oculta para optimizarlo.

3.2.2. Algoritmos de clusterización

Estos algoritmos se utilizan para agrupar datos en clusters cumpliendo las siguientes reglas [22]:

1. Todos los elementos de un cluster deben ser tan similares entre ellos como sea posible.
2. Elementos de dos clusters diferentes deben ser tan diferentes como sea posible.
3. La medición de similaritud debe estar clara y ser lo más práctica e intuitiva posible.

K-Means

El algoritmo K-Means [23] divide una serie de puntos en n dimensiones a lo largo de un numero de clusters predefinido (K). El objetivo es minimizar el error suma de cuadrados dentro de cada cluster. Se ha implementado en WEKA, utilizando la distancia euclídea, una versión modificada llamada KValid [24], que incluye el cálculo del índice de Silueta para estudiar cuál es el número óptimo de clusters.

Expectation Maximization

Cuando el set de datos no tiene variables numéricas (o no tiene sentido calcular distancia euclídea entre 0 y 1 porque es una variable binaria) una solución es utilizar el algoritmo EM. Este algoritmo [25] calcula iterativamente la estimación de máxima similitud y asigna a cada elemento su probabilidad de pertenecer a un cluster.

Se ha utilizado la implementación de Weka para un número indeterminado de clusters, dónde el número óptimo de clusters se obtiene mediante validación cruzada. La validación cruzada (10 fold CV) funciona de la siguiente manera: se divide toda la muestra en 10 grupos, dónde el 90 % es training set y el 10 % test set. Se entrenan los clasificadores en cada grupo y luego se toman las medias de todos para obtener el resultado final.

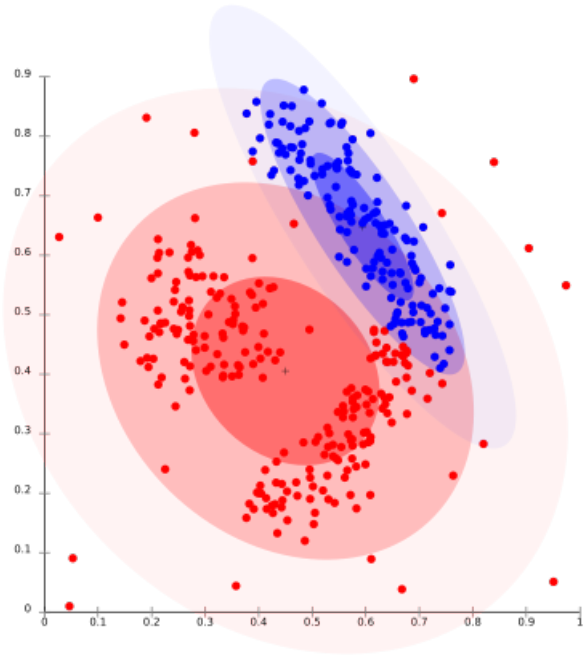


Figura 3.4: Ejemplo de EM en clustering

3.2.3. Algoritmos de clasificación

Los algoritmos de clasificación hacen que el ordenador aprenda de una serie de datos y tras ese aprendizaje pueda decidir si la siguiente observación pertenece a una clase u otra. En este caso se ha utilizado después del proceso de clustering, se ha entrenado asignandole a cada observación su cluster, en un 70 % de la muestra, y luego probando el 30 % restante.

C4.5

Se ha utilizado el algoritmo C4.5, desarrollado por Ross Quinlan, que crea un árbol de decisión a partir de los datos de entrada. Para su implementación se ha utilizado el árbol J48 en Weka.

3.3. Exponente de Lyapunov

Adicionalmente, se ha procedido a calcular los exponentes de Lyapunov [26] para la serie temporal de la demanda. Esta variable mide el grado de separación de la muestra a lo largo de la serie. Se define como:

$$\lambda(x_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln |f(x_i)'|$$

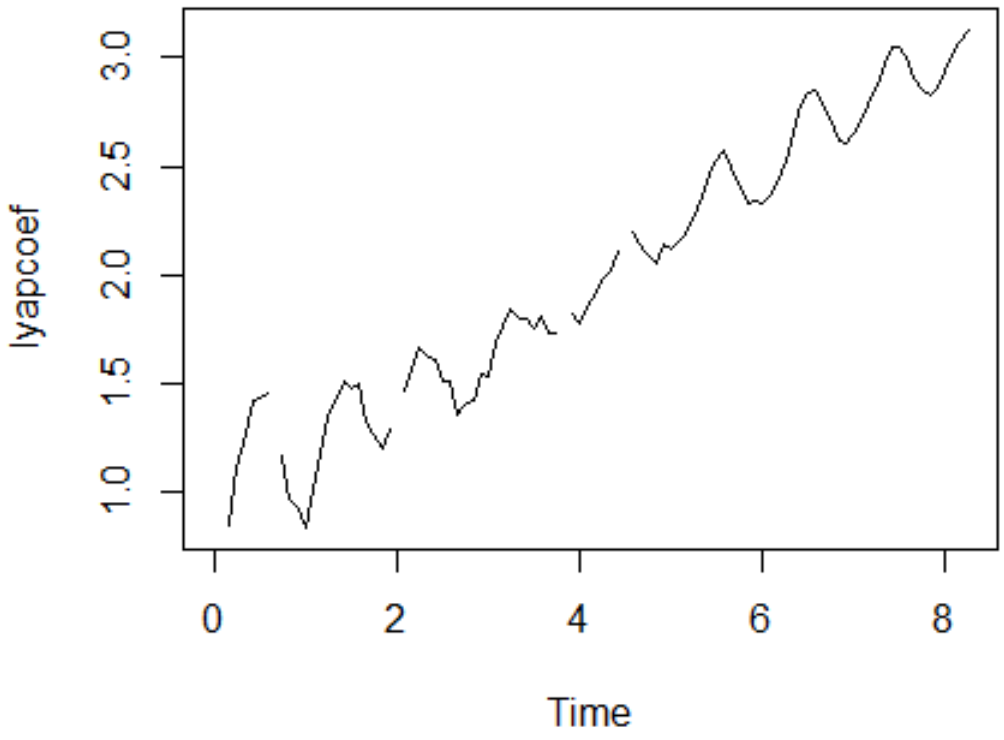


Figura 3.5: Exponente de Lyapunov para demanda 2000 - 2018 (x: años)

Se ha calculado en R, utilizando la librería TSeriesChaos de Antonio Fabio Di Narzo. Se observa una tendencia creciente, prácticamente lineal, por lo que se deduce que la demanda se comporta de manera caótica, con lo cual, por desgracia, no podemos asegurar la significación estadística de este apartado.

4

Evolución económica y social del turismo

Tras haber estudiado los algoritmos a utilizar, en esta sección ya analizamos el desarrollo de las variables económicas comentadas anteriormente. Este análisis se hace desde dos ángulos: un primer análisis predictivo de la demanda, tanto global como en detalle por Comunidades Autónomas y por país de orígen.

El otro foco es estudiar la clusterización de estas variables, utilizando tanto los datos de entrada de turistas en frontera como los datos de gasto, de manera que podamos realizar un análisis geográfico y ver como cambia el peso económico del turismo en España por países.

4.1. Demanda

Demanda total

Para empezar se ha analizado la demanda total en numero de turistas. Esto ya es un problema complejo en sí mismo, se pueden llegar a utilizar modelos muy detallados que relativizan el efecto de la crisis económica de 2008, que eliminan datos muy extremos por encima o por debajo de la media (aunque se pierde información relevante) o que dividen el problema en temporada alta y temporada baja [27].

Para este proyecto se han utilizado dos modelos diferentes: regresión lineal y redes neuronales (preceptrón multicapa). Para evaluar el error se utiliza MAE, MAPE y RMSE, siendo el más importante el MAPE.

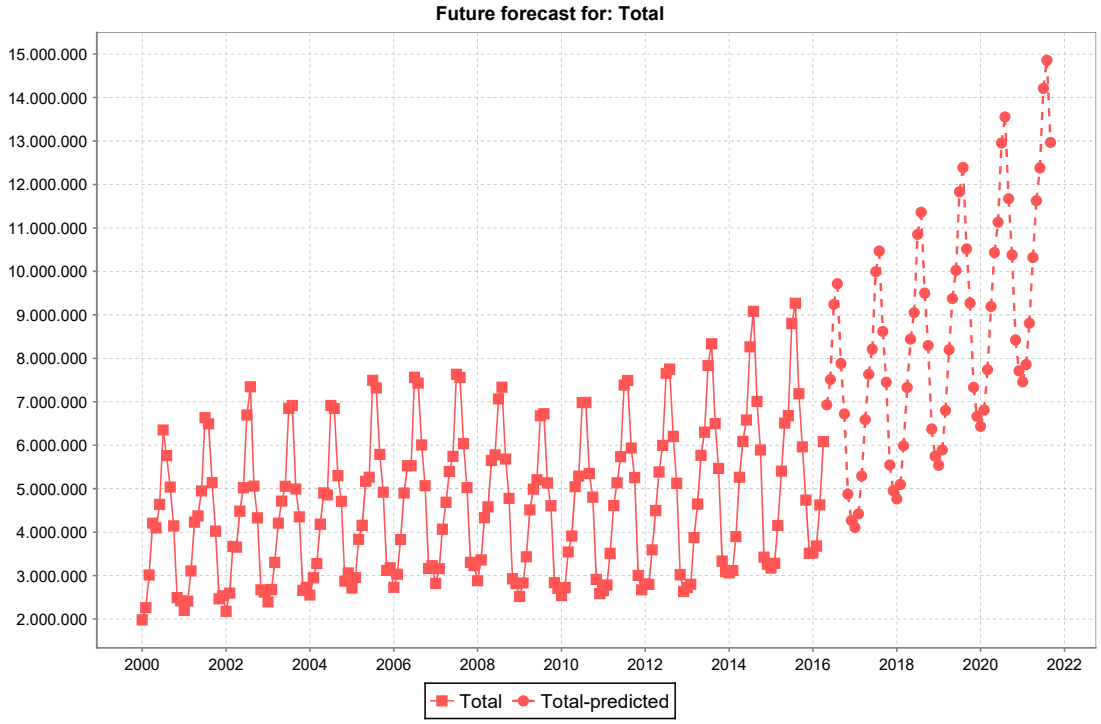


Figura 4.1: Predicción utilizando regresión lineal

El modelo basado en regresión lineal tiene menor error que el basado en redes neuronales, pero hace predicciones demasiado optimistas a partir de un año de horizonte temporal. Este modelo se asemejaría a un escenario donde la economía mundial continúa siendo boyante, los avances tecnológicos en transporte y aplicaciones al turismo continúan sin parar y la tendencia de viaje mundial sigue siendo favorable hacia España. Tiene un MAPE del 8 % para el conjunto de pruebas, y superior al 20 % conforme avanzan las predicciones.

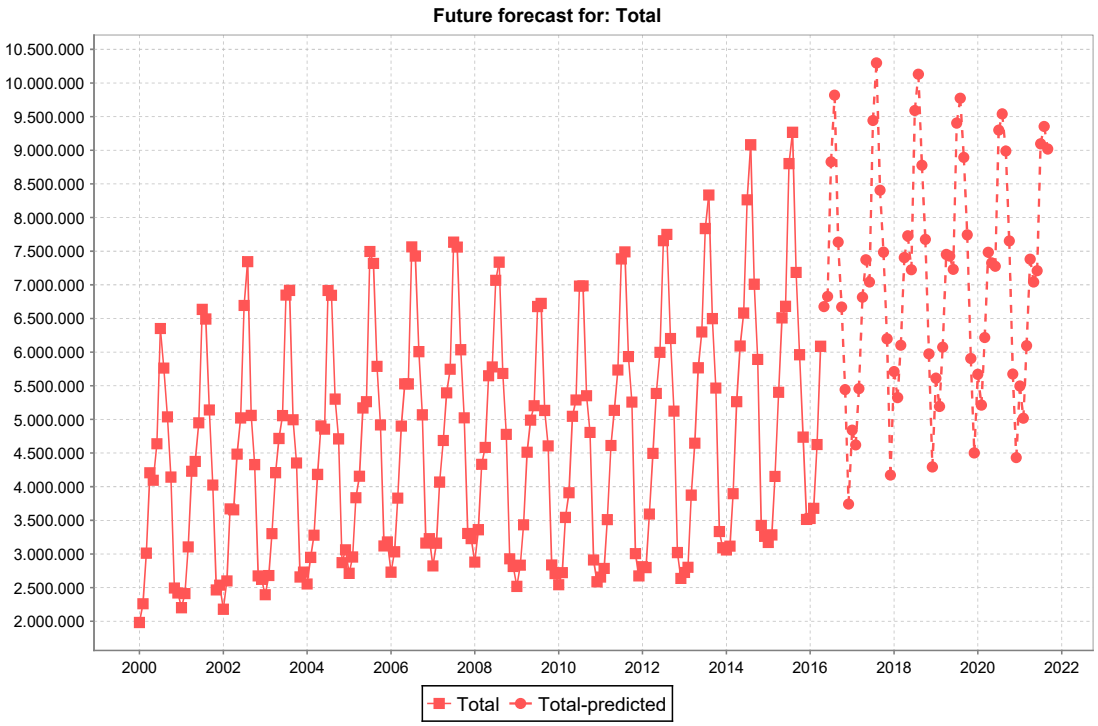


Figura 4.2: Predicción utilizando preceptrón multicapa

El modelo basado en redes neuronales tiene mayor error en el test set porque da mayor peso al efecto de la crisis económica. Este modelo predice un escenario menos favorable, donde las tendencias alcistas empiezan a desaparecer y se otea una nueva crisis económica. Tiene un MAPE del 14 %.

4.1.1. Por Comunidad Autónoma

Podemos hacer este análisis a 1, 2 y 3 años por Comunidades Autónoma (Andalucia, Baleares, Canarias, Cataluña, Valencia, Madrid y otras) para ver un mapa de porcentajes con el peso de cada una en la demanda global. Se ha utilizado el algoritmo SMOReg, con 80 % de set de entrenamiento, ya que es el que menor error producía para cada serie temporal.

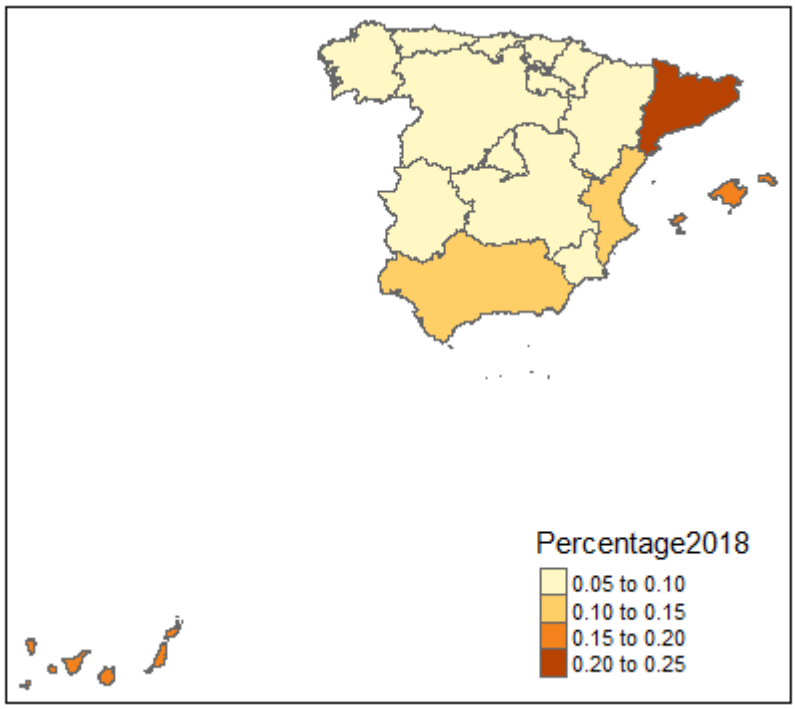


Figura 4.3: Proyección para 2018

La variación que observamos entre los distintos años es principalmente un gran aumento de turistas en Canarias, con un cierto aumento también en Cataluña, mientras que el resto de comunidades crecen ligeramente (con Madrid bajando en algunas franjas temporales).

Una posible explicación a este fenómeno es el gran interés turístico del que goza Canarias, con cada vez más conexiones desde cualquier parte del mundo. Por otro lado, Cataluña ha realizado una labor de promoción enorme en los últimos años y se ve en la tendencia (en estos datos todavía no están contabilizados los meses desde la crisis soberanista, quizá se equilibre la tendencia más a la baja en el futuro).

4.1.2. Por país de orígen

Otro punto que este estudio pretende abordar es la distribución de turistas por país de orígen. Se han obtenido las series temporales para cada país y se han analizado con Weka para obtener la predicción hasta 2023. De aquí se detectan distintos patrones en la serie temporal en función del país de origen:

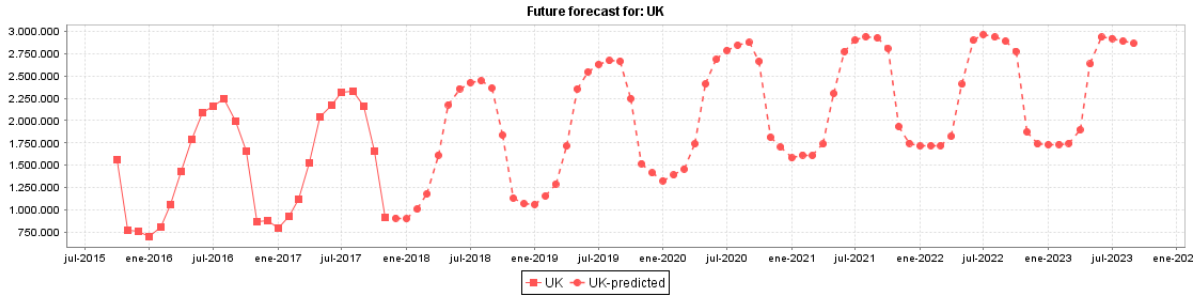


Figura 4.4: Resultado regresión para serie temporal de Reino Unido

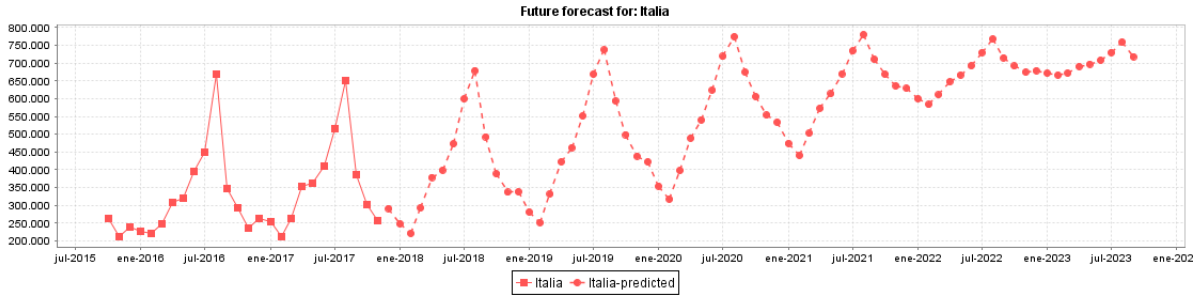


Figura 4.5: Resultado regresión para serie temporal de Italia

Ambas series son cíclicas pero la serie de Reino Unido tiene un comportamiento más sinusoidal y la serie de Italia un comportamiento más triangular. La mayoría de algoritmos de Weka daban resultados completamente fuera de la realidad para este tipo de regresión (más de 60 ticks-meses a futuro) por lo que para las series de este apartado se ha utilizado el algoritmo Preceptrón Multicapa, variando los parámetros (learning rate y momentum) de manera individual para cada una.

Además, otro problema añadido de series como la de Italia es que ha sido imposible encontrar la parametrización que mida con precisión tanto temporada alta como temporada baja, por lo que al final se ha optado por hacer como en el articulo citado anteriormente y estudiar el crecimiento en temporada alta (concretamente, para los mapas, se ha utilizado como punto de referencia Julio).

Por otro lado, el estudio de la forma de las series ya nos está proporcionando información sobre el turismo de un país: países como Reino Unido tienen turistas prácticamente todos los meses del año, por lo que fomentan tanto turismo de playa como turismo cultural, mientras que en países como Italia prima venir en un momento determinado (otras posibles razones pueden ser el sistema de selección de vacaciones en un país y en otro).

Utilizando el resultado de Weka se ha creado un mapa con las tasas de crecimiento para 2021, y otro para 2023. Se han tomado 3 y 5 años como referencias a corto-medio plazo, y se pretende estudiar la evolución de la temporada alta únicamente.

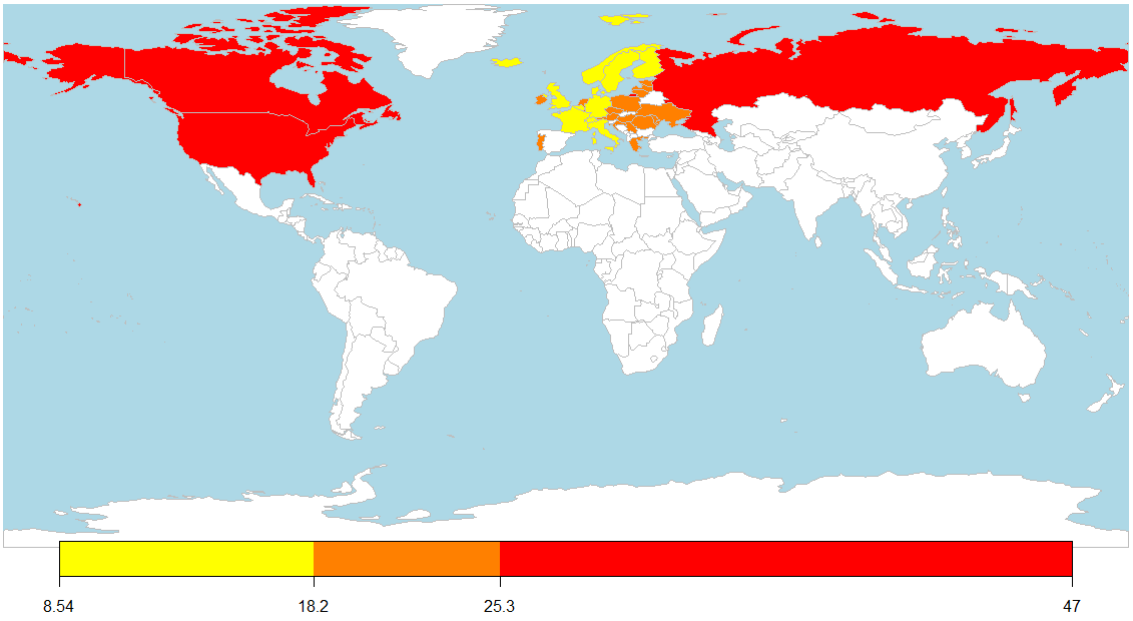


Figura 4.6: Porcentaje de crecimiento por países en 2021 respecto a 2018

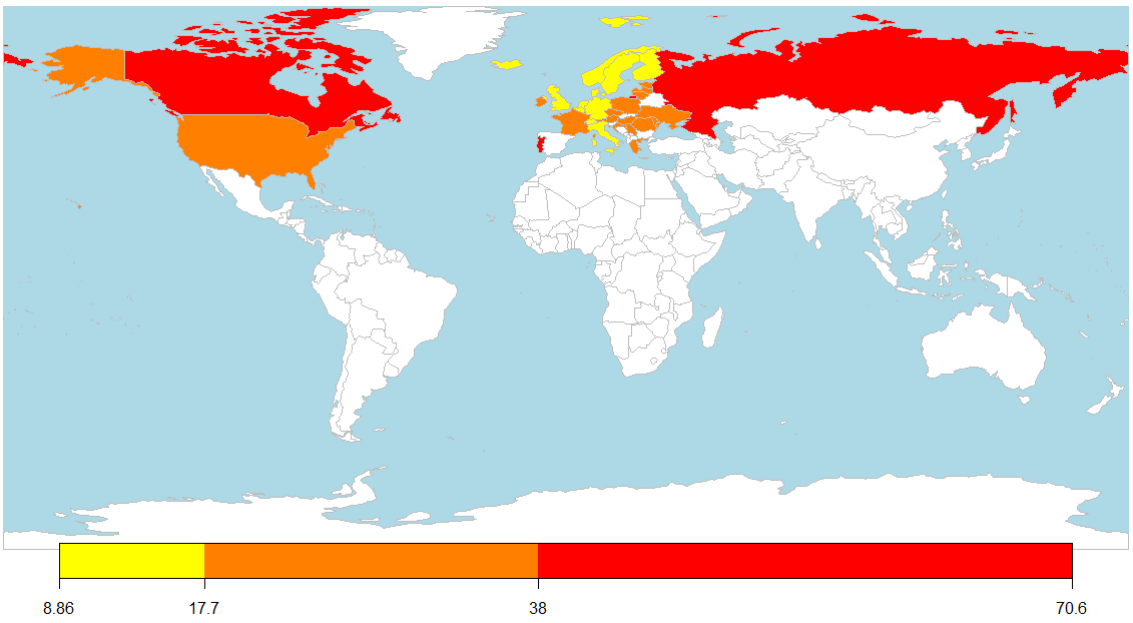


Figura 4.7: Porcentaje de crecimiento por países en 2023 respecto a 2018

Para los países en blanco no se disponía de la información suficiente cómo para hacer un modelo de regresión con cierta precisión.

La primera conclusión que salta a la vista al ver ambos mapas es que hay una gran desigualdad entre los crecimientos previstos para la mayoría de países en Europa respecto al resto del mundo. Tanto para 2021 como para 2023 hay tres bloques: crecimiento bajo, normal y alto.

En 2021 el crecimiento bajo está en nuestros países vecinos, centroeuropa y los países nórdicos. El medio contiene el resto de europa y el crecimiento alto está en Rusia y en América. Para 2023, sin embargo, EEUU ralentiza su crecimiento mientras que países en Europa como Irlanda, Francia o Portugal crecen más. Observando las gráficas individualmente este comportamiento se da porque el crecimiento en EEUU, por ejemplo, aparece como un pico y luego se relaja, mientras que para Portugal es uniforme en el tiempo.

Se adjunta también como información adicional que complementa a los gráficos de crecimiento el mapa con la previsión de número total de turistas por país en 2023. Observamos que hay una relación inversa entre el crecimiento y el número de turistas que tiene un país. Esto puede deberse a que conforme las primeras oleadas de turistas visitan España, el efecto llamada hace que cada vez vengan más. Otra posible explicación, especialmente para Rusia o los países del este de Europa, es que las mejoras en el transporte aéreo les benefician especialmente a ellos respecto a nuestros países vecinos, que no notan tanto la diferencia. Cabe notar la excepción de Francia, que pese a estar en más de 2 millones de turistas, experimenta un crecimiento del 20 al 40 por ciento para 2023.

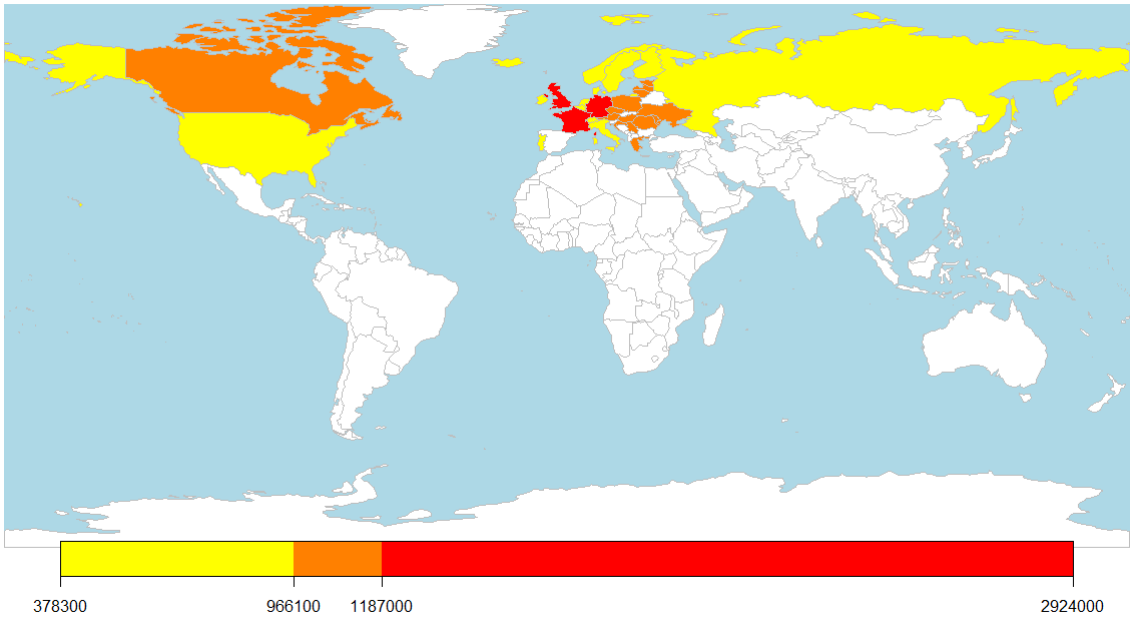
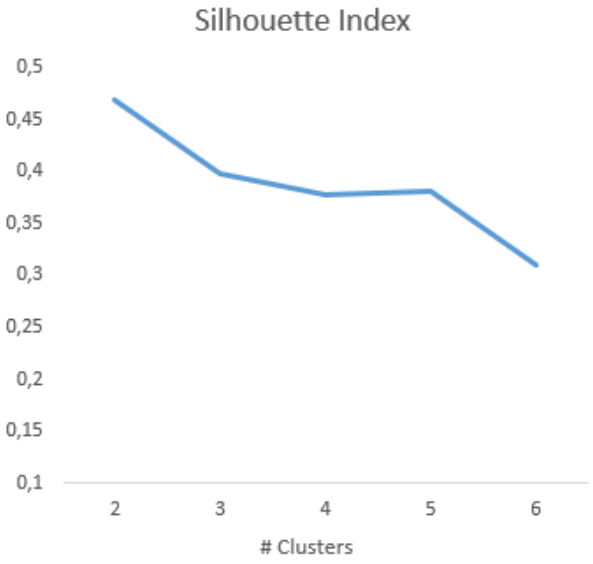


Figura 4.8: Total de turistas previstos por país para Julio 2023

4.2. Segmentación de turistas internacionales

4.2.1. Frontur

Para este apartado se han utilizado los datos (un 30 % tomado al azar de toda la muestra) de Frontur [28] de 2013, 2014 y 2015 para entrenar el modelo. Se ha optado por utilizar 5 clusters tras analizar los distintos resultados de las métricas Silhouette Index y Sum of squared errors. Con 2 clusters se obtiene el mejor resultado de SI, pero con 5 clusters se reduce mucho el SSE manteniendo un SI adecuado.



(a) Valores Silhouete Index

(b) Valores Sum of Squared Errors

La siguiente imagen ilustra el porcentaje de turistas en cada cluster:

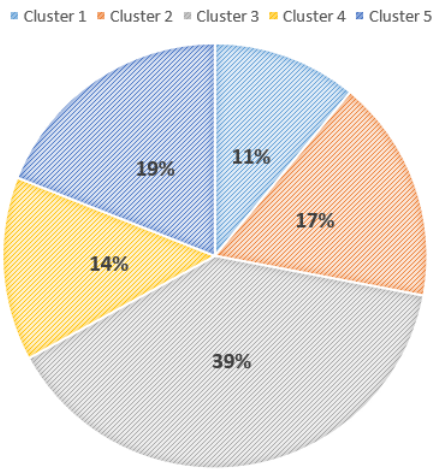


Figura 4.10: % turistas por Cluster

Este gráfico indica que el cluster 3 representa el perfil de turista dominante entre toda la muestra, con un 39 % (aproximadamente 100000 turistas seleccionados). Para estudiar en detalle cada cluster se ha utilizado la interfaz gráfica de WEKA, que permite mostrar gráficamente la distribución de cada cluster para cada variable (se adjunta imagen de ejemplo, ver Anexo I para todas las variables)

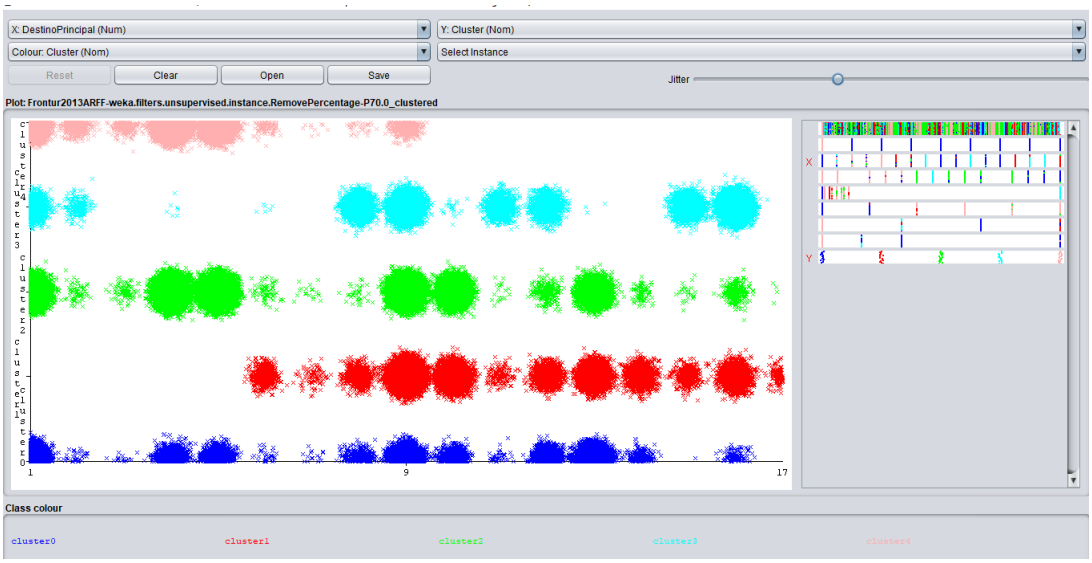


Figura 4.11: Distribución de comunidad autónoma de destino por cluster

Pasamos a caracterizar cada cluster en función de los indicadores:

- Cluster 1: Define uno de los dos clusters de excursionistas (no se alojan durante su viaje). Este grupo viaja a Andalucía, Baleares, Canarias, Cataluña y Madrid mayoritariamente. Otro punto diferenciador es que son principalmente excursionistas de Francia, Italia, Reino Unido y América. Mayormente se quedan durante ninguna hasta un par de noches. Además, no hay preferencia mayoritaria entre entrar a España por tierra, mar o aire.

- Cluster 2: Cluster de turistas no residentes en España, principalmente de Alemania, Bélgica, Francia, Holanda y Reino Unido (podríamos caracterizar este cluster como turista centroeuropeo). Se divide entre una mitad que no se aloja o se alojan en viviendas particulares. Se quedan más de tres días, siendo la media una semana aproximadamente. Utilizan carretera y avión para entrar a España y visitan especialmente Cataluña, Comunidad Valenciana, Madrid, Galicia y País Vasco.

- Cluster 3: Segundo cluster de turistas no residentes. Visitan Andalucía, Baleares, Canarias, Cataluña, Valencia y Madrid (podríamos caracterizarlo como turistas de playa + Madrid). Engloba muchos países, prácticamente todo el mundo salvo el bloque centroeuropeo ya descrito en el cluster 2 (y salvo España). Se quedan de 4 a 7 noches de media, utilizan en misma proporción todo tipo de alojamientos (desde hoteles a vivienda particular), y utilizan avión para desplazarse.

- Cluster 4: Segundo cluster de excursionistas. Utilizan carretera para entrar a España, y realizan excursiones en Andalucía, Castilla La Mancha, Cataluña, Extremadura, Galicia, Navarra y País Vasco, por lo que podríamos definir este cluster como excursionistas de montaña (turismo rural). Vienen de Francia, Portugal y países minoritarios de Europa.
- Cluster 5: Tercer y último cluster de turistas. Visitan Andalucía, Aragón, Asturias, Baleares, Canarias y Cataluña, y vienen de nuestros países vecinos en Europa. Se quedan de media más de una semana, y utilizan hoteles, campings y resorts principalmente.

Una vez analizados los clusters, se ha procedido a añadir como nueva variable a qué cluster corresponde cada elemento de la muestra. Después se han utilizado estos datos para construir un modelo de clasificación, y se ha probado contra todos los datos para 2013, 2014 y 2015. El objetivo es determinar si se puede predecir con precisión el cluster al que pertenecen todos los turistas (y los nuevos que lleguen en el futuro).

Para ello, en Weka se ha utilizado AddCluster, y luego el algoritmo de clasificación J48 (con el 30 % de datos de training set, y el 70 % de prueba). Obtenemos un RRSE (Error relativo al cuadrado) del 2 %, con lo que se puede afirmar que el modelo entrenado para estos 5 clusters organiza efectivamente los datos.

4.2.2. Egatur

Hemos comprobado que segmentar los turistas internacionales por preferencias tiene sentido, el problema ahora es que Frontur no tiene información en detalle sobre estos turistas (sus gustos de ocio, en que gastan el dinero, etcétera). En esta sección se analizará la estadística Egatur (Encuesta sobre el gasto turístico) del INE, con la intención de profundizar en la segmentación y obtener información aprovechable por gobiernos o empresas.

Se ha utilizado la encuestra del 2013, con una muestra de 300000 familias, aproximadamente. Tras el preprocesado y el análisis con Weka, se obtienen 11 clusters, de los cuales tras la fase de testeo 2 desaparecen. Se adjunta como gráfico la distribución de la muestra por cluster.

A continuación se describe el detalle de cada cluster:

- El cluster 0 contiene el 1 % de la muestra. Principalmente de Alemania, Irlanda, Italia, Reino Unido y los países Nórdicos. Están, con 3200€, en el tercer lugar en cuánto a gasto por persona. Los motivos principales de viaje son negocios y ocio de playa. Realizan actividades culturales (el 70 %), y el 50 % acude a restaurantes. El alquiler de coches y la visita de discotecas es marginal. Aproximadamente el 35 % hace excursiones.
- El cluster 1 contiene la mayoría de la muestra, el 34 %. La media de gasto por persona son 2000€. Este cluster está formado por turistas que contratan paquete turístico.

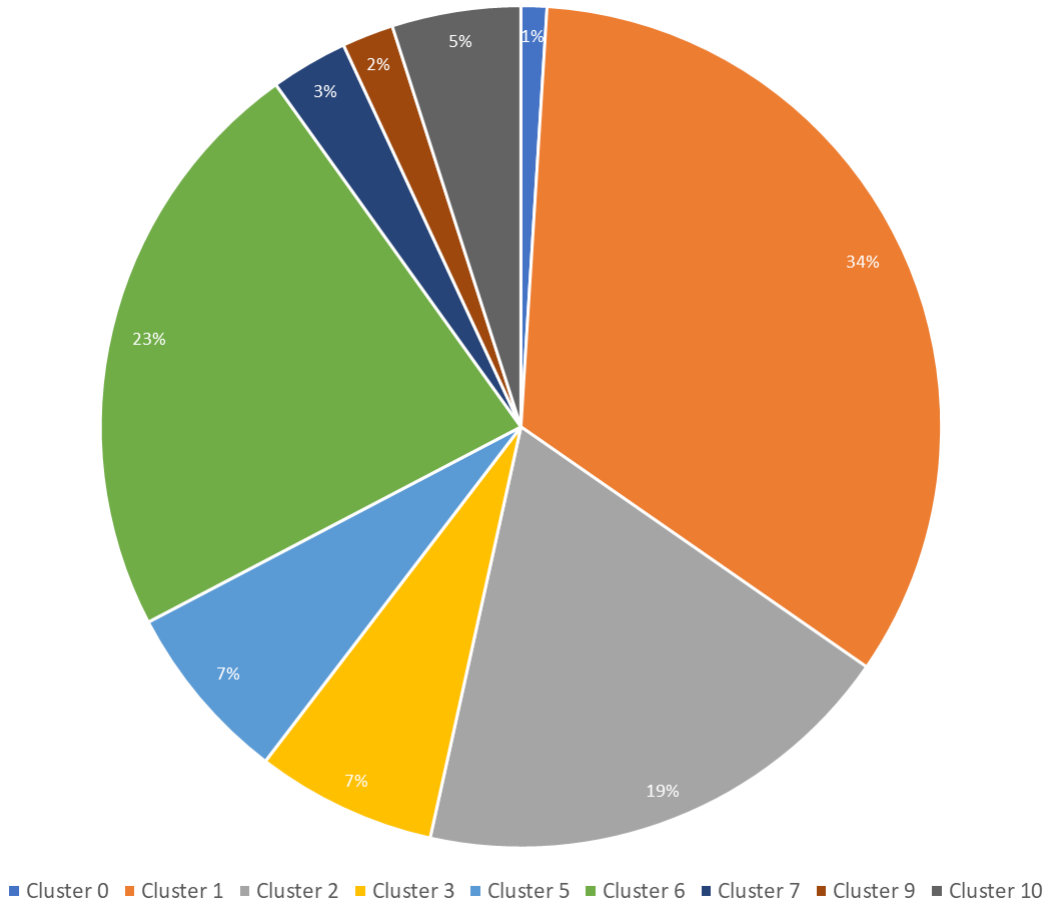


Figura 4.12: % turistas por Cluster (encuesta de gasto)

Contiene tanto turistas que van a la playa como turistas que vienen a visitar a amigos o por negocios. No se aprecian más puntos de segmentación notables.

- El cluster 2 contiene el 19 % de la muestra. El gasto medio por persona es de 400€, el menor de todos. Este cluster comprende todos los turístas de los que no se ha podido obtener más información (indicador: no procede).
- El cluster 3 contiene el 7 % de la muestra. 2100€ es el gasto por persona. Contiene tanto turismo de playa como turismo deportivo (junto con el cluster 7, son los dos únicos clusters con deporte), en este caso gratuíto. Realizan principalmente senderismo, deportes náuticos y de aventura, y otros deportes (futbol, volleyball). Además, el 30 % visita espectáculos deportivos.
- El cluster 5 contiene también el 7 % de la muestra. El gasto de media son 2700€ por persona. Gran concentración de gente que viaja sola, después con amigos o con pareja. Este cluster visita discotecas y clubs (65 %), y espectáculos deportivos (50 %). Lo componen principalmente EEUU, UK y resto de América, junto con Europa. La mayoría contrata paquete turístico.
- El cluster 6 contiene el 23 % de la muestra, con un gasto por persona de 3150€. Contiene principalmente turismo cultural y de playa. Esta formado por gente que

viaja sola o en pareja. Están interesados en excursiones, alquilan coche y vienen con paquete turístico (concentración del 60 % en cada una).

- El cluster 7 contiene el 3 % de la muestra, que gasta 4600€ de media. Es el segundo cluster que realiza deportes, en este caso tanto gratuítos como de pago. Se compone principalmente de turistas Alemanes, Franceses, UK y de países minoritarios en Europa. Es turismo de campo y playa. Con concentraciones cercanas al 70 % se realizan deportes cómo esquí, golf, caza, o barco. Visitan espectáculos deportivos.

- El cluster 9 contiene el 2 % de la muestra. El gasto por persona es de 1500€. Son turistas de EEUU, resto de América y resto del mundo. Cómo el cluster 2, comprende turistas de los que no se ha podido obtener más información.

- El cluster 10 contiene el 5 % de la muestra. La media de gasto por persona es de 5000€, situándolo en la primera posición. Contiene turistas de todo el mundo salvo España, con una especial concentración de turistas de Reino Unido y los países Nórdicos. El motivo principal es el turismo de playa, con una pequeña concentración en turismo cultural. El 70 % realizan actividades culturales, el 50 % van a discotecas y clubs, el 60 % realiza excursiones. Sólo el 30 % alquila coche. El 60 % compra productos típicos. Por último, en su mayoría viajan o con su pareja, o con pareja e hijos, o con amigos.

5
Conclusiones

5.1. Conclusiones

Con este trabajo se ha aprendido que se pueden aplicar técnicas de aprendizaje automático satisfactoriamente sobre datos turísticos, obteniendo conclusiones interesantes. Esto demuestra que empresas y organizaciones deberían destinar cada vez más esfuerzos a este área, y utilizarlo para su toma de decisiones.

En el apartado de predicción de demanda, se concluye que de aquí a 3 años la tendencia sigue al alza, y a horizontes mayores quizá excesivamente al alza (ya que es imposible prevér que se van a mantener estas tasas de crecimiento). Además se concluye que los países que menos turistas traen a España actualmente tienen la mayor proyección de aquí a 5 años. Esto podría abrir oportunidades de negocio a empresas que se enfoquen en ofrecer actividades turísticas a rusos o canadienses, por ejemplo.

En el apartado de segmentación se observa que hay varios grupos diferenciados, especialmente en gasto, donde pequeños porcentajes de la muestra son los que tienen el mayor índice de gasto, y los que realizan actividades especiales como deportes de pago o contratación de paquete turístico. Por el lado de la demanda se observa que turistas de determinados países tienden a permanecer más días en España frente a otros.

Con esta información también aparecen oportunidades de negocio, por ejemplo, usando como referencia los tres clusters con mayor gasto por persona, se puede deducir que el mayor beneficio para una empresa que ofrezca paquetes turísticos es añadir la opción de realizar deportes como golf o navegación, alquiler de coche y actividades culturales.

Además se prueba que es absolutamente necesario para cualquier persona que quiera dedicarse al turismo aprender inglés y francés, ya que la masa actual de turistas contiene una gran concentración de Reino Unido y Francia, y a futuro se ha estudiado que Estados Unidos y Canadá tienen mucho crecimiento.

Por último, como conclusión personal, he aprendido a utilizar la librería Weka, a programar en el lenguaje R, y sobre cómo abordar un problema de aprendizaje automático: tanto la metodología, como qué posibles algoritmos hay, su implementación y la muestra de resultados.

5.2. Retos futuros

Hay una serie de puntos en los que se puede ahondar de cara a futuros trabajos, se destacan aquí los tres más importantes.

En primer lugar, este trabajo se queda en la punta del iceberg dentro de la profundidad y complejidad de esta industria. Siguientes trabajos podrían abordar puramente el tema de la sostenibilidad, o especializarse en un subsector concreto (turismo en hoteles, turismo de un determinado país).

En segundo lugar, y conectando con el primer punto, se reafirma como algo clave el conseguir nuevas fuentes de datos. Para obtener análisis con profundidad y con mucho valor añadido, es muy útil probar diferentes fuentes y cruzar los resultados. Aquí resaltamos el conseguir datos transaccionales por parte de algún banco, datos telefónicos o tener acceso a encuestas especializadas del sector.

Por último, el uso de plataformas más avanzadas que Weka puede permitir mejores resultados (por ejemplo, en la clusterización). Para futuros trabajos, hacer especial énfasis en el tratamiento de los datos, las variables y algoritmos a elegir, y su implementación, mejora el resultado. Como ejemplo estudiar sólo datos de temporada alta ya ha dado buenos resultados en problemas como el turismo en Tailandia.

Bibliografía

[1] «Cuenta Satélite del Turismo de España. Base 2010 Serie 2010-2015». En: (2015). URL: http://www.ine.es/en/prensa/np1015_en.pdf.

[2] «El turismo, motor del crecimiento y de la recuperación de la economía española». En: (2015). URL: https://ebuah.uah.es/dspace/bitstream/handle/10017/21517/turismo_cuadrado_IAESDT_2015_N04.pdf?sequence=1.

[3] «España sella 2017 como segunda potencia turística mundial superando a EEUU y solo por detrás de Francia». En: (2018). URL: http://www.europapress.es/turismo/nacional/noticia-espana-bate-record-82-millones-turistas-internacionales-2017-superaria-eeuu-20180110211537.html.

[4] «España lidera de nuevo el índie de competitividad turística mundial». En: (2016). URL: https://www.hosteltur.com/121402_espana-lidera-nuevo-indice-competitividad-turistica-mundial.html.

[5] «Tendencias y Evolución del Turismo en España Expectativas 2017». En: (2017). URL: https://perspectivas.deloitte.com/expectativas-turismo-2017.

[6] «MEASURING TOURISM'S IMPACT A PILOT STUDY IN CYPRUS». En: (2016). URL: https://s3-eu-west-1.amazonaws.com/travelfoundation/wp-content/uploads/2016/11/17163605/Tourisms_Impact_quality_no_bleed.pdf.

[7] «The Travel and Tourism Competitiveness Report 2017». En: (2017). URL: https://www.weforum.org/reports/the-travel-tourism-competitiveness-report-2017.

[8] «On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers». En: (1927).

[9] «On periodicity». En: (1925).

[10] «Time Series Analysis». En: (1970).

[11] «Recurrent Neural Networks and Robust Time Series Prediction». En: (1994).

[12] «BBVAturismo: An analysis of tourism in Spain through spending». En: (2016). URL: https://www.bbva.com/en/bbvaturismo-analysis-tourism-spain-spending/.

[13] Shu-Hsien Liao, Pei-Hui Chu y Pei-Yuan Hsiao. «Data mining techniques and applications – A decade review from 2000 to 2011». En: *Expert Systems with Applications* 39.12 (2012), págs. 11303 -11311. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2012.02.063. URL: http://www.sciencedirect.com/science/article/pii/S0957417412003077.

[14] Freek Stulp y Olivier Sigaud. «Many regression algorithms, one unified model: A review». En: *Neural Networks* 69 (2015), págs. 60 -79. ISSN: 0893-6080. DOI: https://doi.org/10.1016/j.neunet.2015.05.005. URL: http://www.sciencedirect.com/science/article/pii/S0893608015001185.

[15] Gianluca Bontempi. «Machine Learning Strategies for Time Series Prediction». En: (2013). URL: http://www.ulb.ac.be/di/map/gbonte/ftp/time_ser.pdf.

[16] John Platt. *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*. Inf. téc. 1998, pág. 21. URL: https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/.

[17] Mathias M. Adankon y Mohamed Cheriet. «Support Vector Machine». En: *Encyclopedia of Biometrics*. Ed. por Stan Z. Li y Anil Jain. Boston, MA: Springer US, 2009, págs. 1303-1308. ISBN: 978-0-387-73003-5. DOI: 10.1007/978-0-387-73003-5_299. URL: https://doi.org/10.1007/978-0-387-73003-5_299.

[18] Yann Lecun y col. «Learning Algorithms For Classification: A Comparison On Handwritten Digit Recognition». En: (1995), págs. 261-276.

[19] Edgar Osuna, Robert Freund y Federico Girosi. «Training Support Vector Machines: an Application to Face Detection». En: (1997), págs. 130-136.

[20] S. K. Shevade y col. «Improvements to the SMO algorithm for SVM regression». En: *IEEE Transactions on Neural Networks* 11.5 (2000), págs. 1188-1193. ISSN: 1045-9227. DOI: 10.1109/72.870050.

[21] M.W Gardner y S.R Dorling. «Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences». En: *Atmospheric Environment* 32.14 (1998), págs. 2627 -2636. ISSN: 1352-2310. DOI: https://doi.org/10.1016/S1352-2310(97)00447-0. URL: http://www.sciencedirect.com/science/article/pii/S1352231097004470.

[22] Dongkuan Xu y Yingjie Tian. «A Comprehensive Survey of Clustering Algorithms». En: *Annals of Data Science* 2.2 (2015), págs. 165-193. ISSN: 2198-5812. DOI: 10.1007/s40745-015-0040-1. URL: https://doi.org/10.1007/s40745-015-0040-1.

[23] JA Hartigan y MA Wong. «Algorithm AS 136: A K-means clustering algorithm». En: *Applied Statistics* (1979), págs. 100-108.

[24] «KValid information». En: (2018). URL: https://github.com/Theldus/KValid.

[25] Xin Jin y Jiawei Han. «Expectation Maximization Clustering». En: *Encyclopedia of Machine Learning*. Ed. por Claude Sammut y Geoffrey I. Webb. Boston, MA: Springer US, 2010, págs. 382-383. ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_289. URL: https://doi.org/10.1007/978-0-387-30164-8_289.

[26] *Calculating the Lyapunov Exponent of a Time Series*. 2014. URL: https://blog.abhranil.net/2014/07/22/calculating-the-lyapunov-exponent-of-a-time-series-with-python-code/.

[27] «International Tourists Arrival to Thailand: Forecasting by Non- Linear Model». En: (2014). URL: https://ac.els-cdn.com/S2212567114006911/1-s2.0-S2212567114006911-main.pdf?_tid=c7e60af1-716a-4856-8c90-ba79703a2679&acdnat=1523249804_a5226265369f3633c8e712d3fdc72c78.

[28] «Estadísticas Frontur y Egatur». En: (2018). URL: www.iet.tourspain.es.